基于用户扩展兴趣的微博推荐方法 *

徐建民，刘明艳，王 苗†

(河北大学 网络空间安全与计算机学院, 河北 保定 071002)

摘 要：为解决微博用户兴趣提取不准确的问题，提出一种基于用户扩展兴趣的微博推荐方法。该方法将用户个体兴趣与关联兴趣结合为用户扩展兴趣进行微博推荐。其中，用户个体兴趣从用户标签、发布微博及交互微博中提取；用户关联兴趣通过用户与其关注用户间的关注关系强度、交互频繁程度和个体兴趣相似度获取。最后，计算用户扩展兴趣与待推荐微博的相似度，对相似度降序排列产生推荐列表。实验结果表明，新方法较传统方法更具有效性和准确性。

关键词：个体兴趣；关联兴趣；扩展兴趣；微博推荐

中图分类号：TP393.02 doi: 10.3969/j.issn.1001-3695.2018.05.0264

Microblog recommendation method based on extended interest of users

Xu Jianming, Liu Mingyan, Wang Miao†

(*School of Cyber Security & Computer*, *Heibei University*, *Baoding Hebei* 071002, *China*)

Abstract: In order to deal with the problem of extracting interest of microblog users inaccurately, this paper proposed a microblog recommendation method based on extended interest of users. This method combined individual interest and associated interest to represent extended interest for recommending microblogs. It extracted individual interest of users from their tags, posted microblogs and interacted microblogs. Then it got associated interest by the strength of following/followed links, interaction frequency and individual interest similarity between users and their followee. Finally, calculate the similarity between extended interest of users and the microblogs to be recommended, it generated the recommendation lists by descending the order of similarity. Experimental results show the method is more effective and precise than the traditional methods.

Key words: individual interest; associated interest; extended interest; microblog recommendation

0 引言

随着新兴社交媒体的流行，微博已经成为人们分享、传播、获取信息的重要平台[1]。用户的爆炸式增长导致微博平台产生的信息呈指数级增长，信息过载问题日益加剧[2]。因此，针对微博用户的潜在需求，推荐其感兴趣的微博变得尤为重要。

实现微博推荐的关键是提取用户兴趣[3]。高明等人[4]利用 LDA 主题模型推断用户发布微博的主题分布获取用户兴趣；王宁宁等人[5]利用 TextRank 排序方法提取用户发布微博中的关键词作为用户兴趣；Zhou 等人[6]通过构建用户标签图表示用户兴趣。上述方法从用户自身信息中提取用户兴趣，取得了一定的推荐效果，但均未考虑用户间的关注关系。理论上，关注行为可直接反映用户的兴趣导向[7]，将其用于用户兴趣的挖掘是可行的。例如，谭晋秀等人[8]使用 K-means 聚类方法对“特别关注”用户发布的微博进行聚类，提取每类微博中的关键词作为用户兴趣；Ma 等人[9]从用户发布的微博中提取关键词，将关键词与用户在微博系统中选择的标签结合构建用户标签矩阵，并利用标签关联度和用户的关注关系更新用户标签矩阵，得到最终的用户兴趣。上述方法仅考虑了用户间静态的关注关系，不能准确衡量用户与其关注用户的关系强度，并且关注用户存在与用户不同的兴趣，导致提取的用户兴趣不准确。微博用户除关注行为外，还常常对喜爱的微博进行点赞、转发和评论等动态的交互行为，将这些行为用于提取用户兴趣，一方面能较准确地反映用户的兴趣，另一方面相对于静态的关注关系能较好地体现用户之间的关联程度。

针对以上问题，本文考虑了用户间的关注关系和交互行为，提出一种基于用户扩展兴趣的微博推荐方法。该方法通过引入开关变量、调和参数，将用户个体兴趣和用户关联兴趣融合得到用户扩展兴趣。用户个体兴趣从用户自身信息中提取，所用信息包括用户标签、用户发布的微博及交互微博；用户关联兴趣通过用户间的关联度和关注用户的个体兴趣计算得到。

收稿日期：2018-05-21；**修回日期：**2018-07-09 **基金项目：**河北省自然科学基金资助项目（2015201142）；国家社科基金后期资助项目（17FTQ002）

作者简介：徐建民（1966-），男，河北邯郸人，教授，博士，主要研究方向为信息检索、社交网络；刘明艳（1993-），女，河北石家庄人，硕士研究生，主要研究方向为信息检索、微博推荐；王苗（1975-），女（通信作者），河北保定人，副教授，硕士，主要研究方向为信息检索（2629101909@qq.com）.

1　用户扩展兴趣

1.1　用户个体兴趣

用户个体兴趣普遍被描述为用户对各个兴趣词的喜好程度[10]，可从用户标签、用户发布的微博及交互微博中提取关键词来表示用户的个体兴趣词，并用关键词的权重表示用户的喜好程度。

1.1.1 用户标签表示

用户依据其所在领域或兴趣爱好给自己添加标签[11]，可以选择添加微博系统中提供的标签，也可以自定义输入。标签的内容包含一些标识兴趣和身份的短语，如“旅行”“音乐迷”“篮球爱好者”等。用户 u_i 的标签 Pt_i 表示为

$$Pt_i = ((t_{i1}:ft_{i1}),(t_{i2}:ft_{i2}),\cdots,(t_{ih}:ft_{ih}))$$

其中：t_{ik} 为用户 u_i 的第 k 个标签，ft_{ik} 为 t_{ik} 的权重。对于同一个用户，认为用户的每个标签同等重要，即 $ft_{ik}=1/h$，h 为用户 u_i 的标签总数。

1.1.2 用户发布的微博表示

微博用户通过发布微博记录日常生活、表达自己的观点，在一定程度上，用户发布的微博可以表明其个体兴趣[12]。而微博是典型的短文本，若使用词频统计关键词，效果并不理想。因此，本文将用户近期发布的所有微博拼接为一个长文本，然后提取关键词。

用户 u_i 近期发布的微博拼接的长文本 Pp_i 表示为

$$Pp_i = ((p_{i1}:fp_{i1}),(p_{i2}:fp_{i2}),\cdots,(p_{il}:fp_{il})),$$

其中：p_{ik} 为从长文本中提取的第 k 个关键词，l 为关键词总数，fp_{ik} 为 p_{ik} 的权重，采用式（1）计算。

$$fp_{ik} = \frac{N(p_{ik})}{\sum_{k=1}^{l} N(p_{ik})} \tag{1}$$

其中：$N(p_{ik})$ 为 p_{ik} 在长文本中出现的次数。

1.1.3 用户的交互微博表示

用户的交互微博指其点赞、转发和评论的微博。不同交互行为可反映用户对微博的喜爱程度，将用户近期点赞、转发和评论的微博分别拼接为三个长文本 Pz_i、Pr_i 和 Pc_i，表示为

$$Pz_i = ((z_{i1}:fz_{i1}),(z_{i2}:fz_{i2}),\cdots,(z_{ia}:fz_{ia}))$$

$$Pr_i = ((r_{i1}:fr_{i1}),(r_{i2}:fr_{i2}),\cdots,(r_{if}:fr_{if}))$$

$$Pc_i = ((c_{i1}:fc_{i1}),(c_{i2}:fc_{i2}),\cdots,(c_{im}:fc_{im}))$$

其中：z_{ik}，r_{ik}，c_{ik} 分别为从 Pz_i，Pr_i 和 Pc_i 中提取的第 k 个关键词，a，f，m 分别为从 Pz_i，Pr_i 和 Pc_i 中提取的关键词总数，fz_{ik}，fr_{ik}，fc_{ik} 分别为 z_{ik}，r_{ik}，c_{ik} 的权重，采用式（1）计算。

将 Pz_i、Pr_i 和 Pc_i 中的关键词进行合并，并赋予不同权重，则用户 u_i 的交互微博 Pb_i 表示为

$$Pb_i = ((b_{i1}:fb_{i1}),(b_{i2}:fb_{i2}),\cdots,(b_{is}:fb_{is}))$$

其中：b_{ik} 为用户 u_i 交互微博中的第 k 个关键词，s 为关键词总数，$0<s\le a+f+m$，fb_{ik} 为 b_{ik} 的权重，采用式（2）计算。

$$fb_{ik} = \alpha_1 \times wl \times fz_{ik} + \alpha_2 \times wr \times fr_{ik} + \alpha_3 \times wc \times fc_{ik} \tag{2}$$

其中：wl，wr，wc 为点赞、转发和评论的权重，$wl+wr+wc=1$。α_1，α_2，α_3 是开关变量。若 b_{ik} 是从用户 u_i 点赞的微博中提取的关键词，则 α_1 取 1，否则取 0；若 b_{ik} 是从用户 u_i 转发的微博中提取的关键词，则 α_2 取 1，否则取 0；若 b_{ik} 是从用户 u_i 评论的微博中提取的关键词，则 α_3 取 1，否则取 0。

1.1.4 用户个体兴趣表示

定义 1　用户个体兴趣。用户个体兴趣是从用户自身信息中挖掘的兴趣，表示为用户的个体兴趣词与其对个体兴趣词的喜好程度所构成的二元组向量：$P_i = ((w_{i1}:fw_{i1}),(w_{i2}:fw_{i2}),\cdots,(w_{ie}:fw_{ie}))$。其中，$w_{ik}$ 为用户 u_i 的个体兴趣词，$w_{ik}\in\{t_{i1},t_{i2},\cdots,t_{ih}\}\cup\{p_{i1},p_{i2},\cdots,p_{il}\}\cup\{b_{i1},b_{i2},\cdots,b_{is}\}$，$fw_{ik}$ 为 w_{ik} 的权重，计算公式如（3）所示，e 为用户 u_i 的个体兴趣词总数，$0<e\le h+l+s$。

$$fw_{ik} = \beta_1 \times ft_{ik} + \beta_2 \times fp_{ik} + \beta_3 \times fb_{ik} \tag{3}$$

其中：β_1，β_2，β_3 是开关变量，分别标志 w_{ik} 是否为用户 u_i 的标签，从发布微博中提取的关键词和从交互微博中提取的关键词。

对 fw_{ik} 归一化处理，最终用户 u_i 的个体兴趣词 w_{ik} 的权重 fw'_{ik} 计算公式如式（4）所示。

$$fw'_{ik} = \frac{fw_{ik}}{\max\{fw_{i1},fw_{i1},\cdots,fw_{ie}\}} \tag{4}$$

1.2　用户关联兴趣

用户关联兴趣受其关注用户的影响，影响程度通过用户与其关注用户的关联度量化。

1.2.1 用户间的关联度

用户 u_i 与用户 u_j 的关联度是彼此关联程度的体现，由两者的关系紧密度和个体兴趣相似度共同决定，计算公式如（5）所示。

$$F_{ij} = \frac{1}{2}(G_{ij}+I_{ij}) \tag{5}$$

其中：G_{ij} 为用户 u_i 与用户 u_j 的关系紧密度，由用户 u_i 与用户 u_j 的关注度和交互度共同决定，计算公式如式（6）所示。I_{ij} 为用户 u_i 与用户 u_j 的个体兴趣相似度，采用余弦相似度公式计算。

$$G_{ij} = \frac{1}{2}(A_{ij}+S_{ij}) \tag{6}$$

其中：A_{ij} 为用户 u_i 与用户 u_j 的关注度，反映了彼此间的关注关系强度，计算公式如式（7）所示。

S_{ij} 为用户 u_i 与用户 u_j 的交互度,表现为用户间点赞、转发和评论的频繁程度,计算公式如式(8)所示。

$$A_{ij}=\begin{cases}0 & ，u_i与u_j互不关注\\0.5 & ，u_i与u_j单向关注\\1 & ，u_i与u_j双向关注\end{cases}\tag{7}$$

直观上，在微博中，三种关注关系下用户间的关注度从小到大的次序为：用户间互不关注的关注度，用户间单向关注的关注度，用户间双向关注的关注度，据此，本文将 A_{ij} 在三种情况下的值分别定为 0，0.5 和 1。

$$S_{ij}=wl\times SL_{ij}+wr\times SR_{ij}+wc\times SC_{ij}\tag{8}$$

其中：SL_{ij}，SR_{ij}，SC_{ij} 分别表示用户 u_i 与用户 u_j 的点赞交互度、转发交互度和评论交互度，计算方法如式（9）～（11）所示。

$$SL_{ij}=\frac{1}{2}\times\left(\frac{NL_{ij}}{\sum_{j=1}^{n_{i1}}NL_{ij}}+\frac{NL_{ji}}{\sum_{i=1}^{n_{j2}}NL_{ji}}\right)\tag{9}$$

$$SR_{ij}=\frac{1}{2}\times\left(\frac{NR_{ij}}{\sum_{j=1}^{n_{i3}}NR_{ij}}+\frac{NR_{ji}}{\sum_{i=1}^{n_{j4}}NR_{ji}}\right)\tag{10}$$

$$SC_{ij}=\frac{1}{2}\times\left(\frac{NC_{ij}}{\sum_{j=1}^{n_{i5}}NC_{ij}}+\frac{NC_{ji}}{\sum_{i=1}^{n_{j6}}NC_{ji}}\right)\tag{11}$$

其中：NL_{ij}，NR_{ij}，NC_{ij} 为用户 u_i 对用户 u_j 点赞、转发和评论的次数，n_{i1}，n_{i3}，n_{i5} 为用户 u_i 点赞、转发和评论的用户总数，n_{j2}，n_{j4}，n_{j6} 分别为用户 u_j 点赞、转发和评论的用户总数。

1.2.2 用户关联兴趣表示

利用 1.2.1 节中的方法计算用户 u_i 与其关注用户的关联度，筛选出与其关联度大于一定阈值的关注用户 $u_1,u_2,\cdots,u_n$，其中用户 u_j 的个体兴趣表示为

$$P_j=\left(\left(w_{j1}:fw'_{j1}\right),\left(w_{j2}:fw'_{j2}\right),\cdots,\left(w_{jsum_j}:fw'_{jsum_j}\right)\right)。$$

定义 2 用户关联兴趣。用户关联兴趣是从与用户关联度大于一定阈值的关注用户中挖掘的兴趣，表示为用户的关联兴趣词与其对关联兴趣词的喜好程度所构成的二元组向量：

$$Q_i=\left(\left(q_{i1}:fq_{i1}\right),\left(q_{i2}:fq_{i2}\right),\cdots,\left(q_{ig}:fq_{ig}\right)\right)$$

其中：q_{ik} 为用户 u_i 的关联兴趣词，$q_{ik}\in\bigcup_{j=1}^{n}\left\{w_{j1},w_{j2},\cdots,w_{jsum_j}\right\}$，$fq_{ik}$ 为 q_{ik} 的权重，计算公式如（12）所示，g 为用户 u_i 的关联兴趣词总数，并且 $0<g\le sum_1+sum_2+\cdots+sum_n$。

$$fq_{ik}=\sum_{j=1}^{n}\gamma F_{ij}\,fw'_{jk}\tag{12}$$

其中：γ 是开关变量，若 q_{ik} 与 w_{jk} 表示同一个词，则 γ 取 1，否则取 0。

对 fq_{ik} 归一化处理，最终用户 u_i 的关联兴趣词 q_{ik} 的权重 fq'_{ik} 计算公式如式（13）所示。

$$fq'_{ik}=\frac{fq_{ik}}{\max\left\{fq_{i1},fq_{i2},\cdots,fq_{ig}\right\}}\tag{13}$$

1.3 用户扩展兴趣表示

定义 3 用户扩展兴趣。用户扩展兴趣是用户个体兴趣和关联兴趣的调和结果，表示为用户的扩展兴趣词与其对扩展兴趣词的喜好程度所构成的二元组向量：

$$T_i=((d_{i1}:fd_{i1}),(d_{i2}:fd_{i2}),\cdots,(d_{iv}:fd_{iv}))$$

其中，d_{ik} 为用户 u_i 的扩展兴趣词，$d_{ik}\in\{w_{i1},w_{i2},\cdots,w_{ie}\}\cup\{q_{i1},q_{i2},\cdots,q_{ig}\}$，$fd_{ik}$ 为 d_{ik} 的权重，计算公式如式（14）所示，v 为用户 u_i 的扩展兴趣词总数，$0<v\le e+g$。

$$fd_{ik}=\gamma_1\lambda\,fw'_{ik}+\gamma_2(1-\lambda)\,fq'_{ik}\tag{14}$$

其中：γ_1，γ_2 为开关变量，γ_1 标志 d_{ik} 是否为用户 u_i 的个体兴趣词，γ_2 标识 d_{ik} 是否为用户 u_i 的关联兴趣词。λ 为调和参数，$\lambda\in[0,1]$，由式（14）可知，λ 取 0 时，$fd_{ik}=fq'_{ik}$，即用户 u_i 的扩展兴趣词的权重与关联兴趣词的权重相等，此时 T_i 表示用户 u_i 的关联兴趣；λ 取 1 时，$fd_{ik}=fw'_{ik}$，即用户 u_i 的扩展兴趣词的权重与个体兴趣词的权重相等，此时 T_i 表示用户 u_i 的个体兴趣词。

2 微博推荐方法

对于新发布的微博，计算用户扩展兴趣与微博的余弦相似度，通过相似度对微博降序排列，将排名 TOP-N 条微博推荐给用户。微博表示为

$$M_t=((m_{t1}:fm_{t1}),(m_{t2}:fm_{t2}),\cdots,(m_{to}:fm_{to}))$$

其中：m_{tk} 为从微博中提取的第 k 个关键词，fm_{tk} 为 m_{tk} 的权重，计算方法如式（1）所示，o 为关键词总数。

本文微博推荐方法（UEI）的具体过程如算法 1 所示。

算法 1 为用户推荐 TOP-N 条微博

输入：用户的标签向量 Pt_i、发布微博向量 Pp_i、交互微博向量 Pb_i，待推荐微博向量 M_t，阈值 δ，用户的关注用户总数 $A(u_i)$，参数 λ。

输出：用户的 TOP-N 推荐列表。

```
for i=1 to n
  计算用户 u_i 的个体兴趣 P_i
for i=1 to n
  for j=1 to A(u_i)
    Q_i=((0:0))
    计算用户 u_i 与 u_j 的关联度 F_ij
    if( F_ij >阈值 δ )
      计算用户 u_j 的个体兴趣 p_j
      用户 u_i 的关联兴趣 Q_i += F_ij × P_j
for i=1 to n
```

for t=1 to K

　用户 u_i 的扩展兴趣 $T_i=\lambda P_i+(1-\lambda)Q_i$

　计算 T_i 与待推荐微博 M_t 的相似度

　为用户 u_i 产生推荐列表 H_i

return　$\{H_1,H_2,\cdots,H_n\}$

3　实验及分析

3.1　实验数据

因为目前没有统一、权威的微博数据集可供使用，本文的实验数据通过爬虫工具进行抽取和采集。首先选取来自电影、游戏、音乐、美食、财经、房产、赛事和汽车八个领域的 40 个微博认证用户作为目标用户，利用雪球采样的爬行策略，顺着目标用户的关注用户链向外爬行扩展一层，最终获取 3087 个微博用户自 2018 年 3 月 22 日到 2018 年 6 月 21 日发布、点赞、转发和评论的微博数据，共计 137 945 条。此外，实验数据还包括这些用户的标签及用户之间的关注关系。为实现最终的微博推荐，搜集了 2018 年 6 月 22 日到 2018 年 6 月 24 日的 300 条热门微博作为待推荐微博。

3.2　评价标准

本文采用平均倒数排名 MRR 和查准率 P 作为评价推荐方法性能的标准。

平均倒数排名 MRR 表示 top-N 推荐列表中第一条正确微博所在位置的倒数均值，MRR 值越高，表示用户感兴趣的微博在推荐列表中的排名越靠前，推荐顺序越合理，计算公式如式（15）所示。

$$MRR=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{rank_i} \tag{15}$$

其中：n 为目标用户总数，$rank_i$ 为用户 u_i 的推荐列表中第一条正确微博所在位置。

查准率 P 为用户感兴趣的微博在推荐列表中所占比例，P 值越高，表示推荐方法的准确率越高。

3.3　参数及阈值设置

1）交互行为权重

点赞、转发和评论的权重采用层次分析法（AHP）中成对比较矩阵和一致性检验的方法确定。对 wl，wr，wc 进行两两比较，得到的判定矩阵如表 1 所示。判定矩阵的最大特征值为 3.0183，对应的特征向量为：$u=(0.1862,0.8527,0.4881)$，将该向量标准化，得到标准化后的向量为 $u^{'}=(0.1219,0.5584,0.3197)$，其每一维的值分别对应点赞、转发和评论的权重，即 $wl=0.1219, wr=0.5584, wc=0.3197$。

表 1 判定矩阵

影响因素	wl	wr	wc
wl	1	1/4	1/3
wr	4	1	2
wc	3	1/2	1

2）关联度阈值

将用户与其关注用户的关联度最小值作为关联度阈值 δ。若用户 u_i 与用户 u_j 是单向关注关系，没有点赞、转发和评论的交互行为，个体兴趣完全不同时，两者的关联度最小。首先，由式（7）（8）分别计算可得上述情况下两者的关注度为 0.5，交互度为 0；然后，将上述结果代入式（6）计算得到用户 u_i 与用户 u_j 的关系紧密度为 0.25；最后，将用户 u_i 与用户 u_j 的关系紧密度 0.25 和个体兴趣相似度 0 代入式（5）可得两者的关联度为 0.125，故将用户间的关联度阈值 δ 设置为 0.125。

3）调和参数

参数 λ 用于调和用户个体兴趣和关联兴趣所占比重。λ 越大，用户个体兴趣所占比重越大；λ 越小，用户关联兴趣所占比重越大。不同的数据集下参数 λ 的取值应该不同，针对本文的实验数据集，通过反复实验来确定参数 λ 的取值。采用人工标注法，对 40 个目标用户感兴趣的微博进行标注，计算 $\lambda=0,0.1,0.2,\cdots,0.9,1$ 时本文的推荐方法在 top-20 推荐结果下的性能指标，结果如图 1 所示。

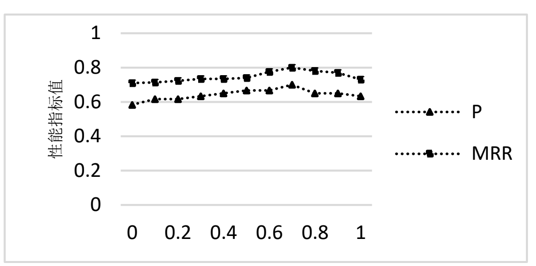


图 1　不同 λ 下 MRR 值和 P 值的对比

从图 1 可以看出，当 $\lambda=0.7$ 时，MRR 值和 P 值均达到最大，故后续实验中将 λ 设置为 0.7。当 $\lambda=0$ 时，用户的扩展兴趣完全取决于关联兴趣；当 $0<\lambda\leq 0.7$ 时，随着 λ 的增大，个体兴趣所占比重增大，关注用户的个体兴趣对用户个体兴趣的干扰逐渐减小，P 值和 MRR 值逐渐增大，即推荐效果逐渐提升；当 $0.7<\lambda\leq 1$ 时，随着 λ 的增大，用户自身数据不足表现越来越明显，导致推荐结果逐渐降低。

3.4　实验对比结果

为验证微博推荐方法的有效性和准确性，在本文数据集上进行两项对比实验，第一项实验用于比较单独基于用户个体兴趣（UPI）、用户关联兴趣（UAI）和两者融合（UEI）的微博推荐方法的性能。考虑到推荐微博数的不同会对推荐的性能造成影响，实验在推荐列表长度为 5、10、15 和 20 时观察推荐方法的性能。图 2 和 3 为三种方法在 top-5，top-10，top-15 和 top-20 推荐结果下的 MRR 值和 P 值。

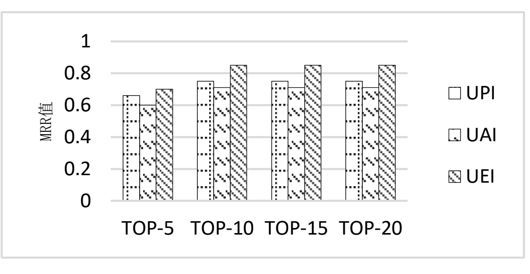


图 2　不同推荐结果下 MRR 值的对比

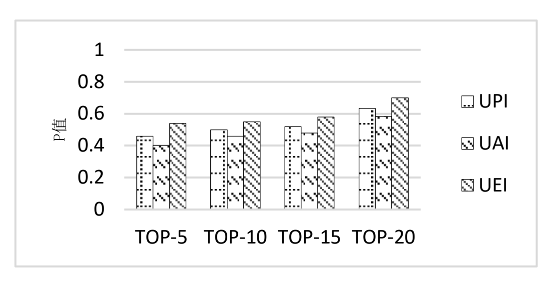


图 3 不同推荐结果下 P 值的对比

从实验结果可以看出：a）采用基于用户个体兴趣（UPI）较基于用户关联兴趣（UAI）的微博推荐方法性能较优，说明用户的关联兴趣只是作为用户个体兴趣的补充，忽略个体兴趣会导致推荐效果不理想；b）本文提出的基于用户扩展兴趣的微博推荐方法（UEI）相对于其他两种方法性能较优，说明将用户个体兴趣和关联兴趣结合可以较为准确地挖掘用户兴趣，缓解提取不活跃用户兴趣困难的问题，进而得到更符合用户兴趣的推荐结果。

第二项实验比较传统的基于内容的微博推荐方法（CB）、文献[13]提出的基于标签的微博信息推荐方法（ITCAUSR）以及本文方法（UEI）的性能。其中，基于内容的推荐方法是通过用户发布的微博建立用户兴趣向量；文献[13]通过用户标签与发布微博构建用户初始标签矩阵，利用标签相关性和用户间的关注关系更新初始矩阵，得到用户兴趣。图 4 和 5 为三种方法在 top-5，top-10，top-15 和 top-20 推荐结果下的 MRR 值和 P 值。

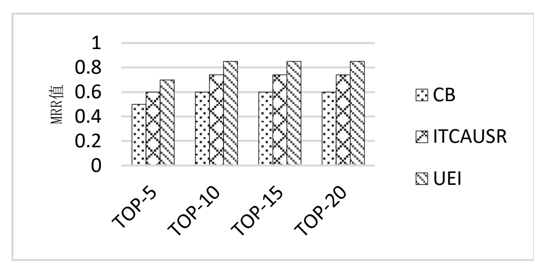


图 4 不同推荐结果下 MRR 值的对比

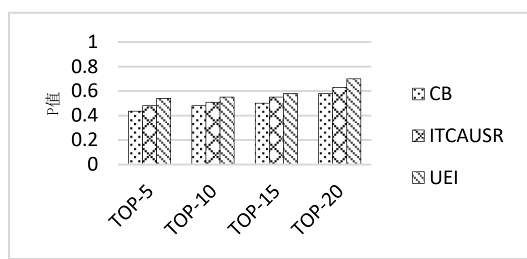


图 5 不同推荐结果下 P 值的对比

由图 3 和 4 的结果可以看出：a）基于标签的微博信息推荐方法（ITCAUSR）较基于内容的微博推荐方法（CB）性能较优，说明在提取用户兴趣时，考虑用户间的关注关系可以使推荐效果得以提升；b）本文提出的基于用户扩展兴趣的微博推荐方法（UEI）相对于基于标签的微博信息推荐方法（ITCAUSR）性能较优，说明考虑用户的交互行为，可以较为准确地提取用户兴趣和区分关注用户与用户的关联程度，达到更好的推荐效果。

4 结束语

本文对微博推荐方法进行了研究，在考虑用户个体兴趣的基础上，结合用户关联兴趣，提出一种基于用户扩展兴趣的微博推荐方法。该方法使提取的用户兴趣更加准确，并在一定程度上解决了提取不活跃用户兴趣困难的问题。实验表明，本文提出的方法相对于以往的 CB 算法和 ITCAUSR 算法有较优的性能。然而，本文仅结合关注用户的兴趣来获取用户扩展兴趣，没有考虑用户的粉丝、与其存在交互行为的非关注用户对其兴趣的影响。因此，确定上述用户的兴趣是否可以作为用户扩展兴趣的一部分将是下一步研究工作中的重点。

参考文献：

[1] Ye Peng, Wang Changbo, Liu Yuhua, *et al.* Visual analysis of micro-blog retweeting using an information diffusion function [J]. Journal of Visualization, 2016, 19 (4): 823-838.

[2] Xu Yan, Zhou Meilin, Han Siyao. Feature representation for microblog followee recommendation in classification framework [C]// Proc of the 7th International Conference on Advanced Computational Intelligence. Piscataway, NJ: IEEE Press, 2015: 318-322.

[3] Zhang Junjie, Lei Yongmei. Improving content recommendation in social streams via interest model [J]. Computer and Information Science, 2015, 566 (1): 57-70.

[4] 高明，金澈清，钱卫宁，等. 面向微博系统的实时个性化推荐 [J]. 计算机学报, 2014, 37 (4): 963-975. (Gao Ming, Jin Cheqing, Qian Weining, *et al.* Real-time and personalized recommendation on microblogging system [J]. Chinese Journal of Computers, 2014, 37 (4): 963-975.)

[5] 王宁宁，鲁燃，王智昊，等. 基于用户标签的微博推荐算法 [J]. 计算机应用研究, 2017, 34 (1): 58-61. (Wang Ningning, Lu Ran, Wang Zhihao, *et al.* Microblog recommendation algorithm based on usrs’ tag [J]. Application Research of Computers, 2017, 34 (1): 58-61.)

[6] Zhou Xianke, Wu Sai, Chun Chun, *et al.* Realtime recommendation for microblog [J]. Information Sciences, 2014, 279: 301-325.

[7] 赵玲，张静. 微博用户行为研究的多维解析 [J]. 情报资料工作, 2013 (5): 65-70. (Zhao Ling, Zhang Jing. Multi-dimensional analysis of microblog behavior research [J]. Information and Documentation, 2013 (5): 65-70.)

[8] 谭晋秀，何跃. 基于 K-means 文本聚类的新浪微博个性化博文推荐研究 [J]. 情报科学, 2016, 34 (4): 74-79. (Tan Jinxiu, He Yue. Study on sina microblog personalized recommendation based on K-means t-ext clustering [J]. Information Science, 2016, 34 (4): 74-79.)

[9] Ma Huifang, Jia Meihuizi, Zhang Di, *et al.* Comb-ining tag correlation and user social relation for microblog recommendation [J]. Information Sciences, 2017, 385-386: 325-337.

[10] 仲兆满，管燕，胡云等. 基于背景和内容的微博用户兴趣挖掘 [J]. 软件学报, 2017, 28 (2): 278-291. (Zhong Zhaoman, Guan Yan, Hu Yun, *et al.*

Mining user interest on microblog based on profile and content [J]. Journal of Software, 2017, 28 (2): 278-291.)

[11] Zhu Peisong, Qian Tieyun, Zhong Ming, *et al.* Inferring users' gender from interests: A tag embedding approach [J]. Neural Information Processing, 2016, 9950: 86-94.

[12] Liu Zhiyuan, Chen Xinxiong, Sun Maosong. Mining the interests of Chinese microblogs via keyword extraction [J]. Frontiers of Computer Science, 2012, 6 (1): 76-87.

[13] 贾美惠子. 基于标签的微博信息推荐技术研究 [D]. 兰州：西北师范大学, 2016. (Jia Meihuizi. Research on microblog recommendation based on tags [D]. Lanzhou: Northwest Normal University, 2016.)